

BALANCED VERTICES IN LABELED ROOTED TREES

MIKLÓS BÓNA

ABSTRACT. In a rooted tree, we call a vertex *balanced* if it is at equal distance from all its descendant leaves. We count balanced vertices in three different tree varieties. For decreasing binary trees, we can prove that the probability that a vertex chosen uniformly at random from the set of all trees of a given size is balanced is monotone decreasing.

INTRODUCTION

Various parameters of many models of random rooted trees are fairly well understood *if they relate to a near-root part of the tree or to global tree structure*. The first group includes the numbers of vertices at given distances from the root, the immediate progeny sizes for vertices near the top, and so on. See [5] for a comprehensive treatment of these results.

Not surprisingly, the technical details of fringe analysis become quite complex as soon as the focus shifts to layers of vertices further away from the leaves. So while there are explicit results on the (limiting) fraction of vertices at a fixed, small distance from the leaves, an asymptotic behavior of this fraction, as a function of the distance, remained an open problem. Recently, Boris Pittel and the present author have studied this family of questions in [2]. Most work in fringe analysis focused on *decreasing binary trees*, which are also called *binary search trees*. We will explain the reason for that. However, in this paper, we will discuss questions that we can successfully investigate for other tree varieties as well.

We call a vertex v of a rooted tree *balanced* if all descending paths from v to a leaf have the same length. In other words, if ℓ is any leaf that is a descendant of v , then the unique path from v to ℓ consists of k edges, where k does not depend on the choice of ℓ . The number k is called the *rank* of v .

1. DECREASING BINARY TREES

A decreasing binary tree on vertex set $[n] = \{1, 2, \dots, n\}$ is a binary plane tree in which every vertex has a smaller label than its parent. This means that the root must have label n , every vertex has at most two children, and that every child v is either a left child or a right child of its parent, even if v is the *only* child of its parent.

Decreasing binary trees on vertex set $[n]$ are in bijection with permutations of $[n]$. In order to see this, let $\pi = \pi_1\pi_2\cdots\pi_n$ be a permutation. The

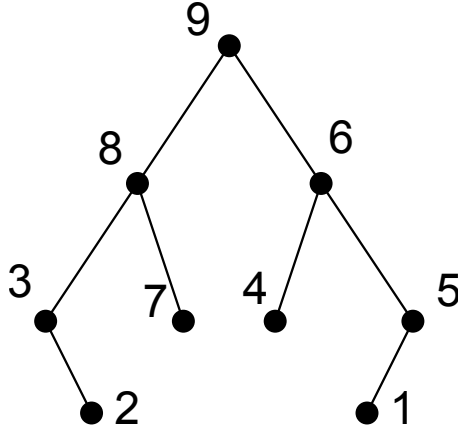


FIGURE 1. The tree $T(\pi)$ for $\pi = 328794615$.

decreasing binary tree of π , which we denote by $T(\pi)$, is defined as follows. The root of $T(\pi)$ is a vertex labeled n , the largest entry of π . If a is the largest entry of π on the left of n , and b is the largest entry of π on the right of n , then the root will have two children, the left one will be labeled a , and the right one will be labeled b . If n is the first (resp. last) entry of π , then the root will have only one child, and that is a left (resp. right) child, and it will necessarily be labeled $n - 1$ as $n - 1$ must be the largest of all remaining elements. Define the rest of $T(\pi)$ recursively, by taking $T(\pi')$ and $T(\pi'')$, where π' and π'' are the substrings of π on the two sides of n , and affixing them to a and b . See Figure 1 for an illustration.

Note that in the tree $T(\pi)$ shown in Figure 1, all vertices are balanced, except 8, 9, and 6. Also note that in any decreasing binary tree, vertex v can be balanced only if all its descendants are balanced.

1.1. Balanced vertices of a fixed rank. Recall that a vertex v is *balanced* if all descending paths from v to a leaf have the same length. That length is the rank of v . Let $a_{n,k}$ be the total number of all balanced vertices of rank k in all decreasing binary trees of size n . Let $A_k(x) = \sum_{n \geq 0} a_{n,k} \frac{x^n}{n!}$, and let $r_{n,k}$ be the total number of such trees on n vertices whose *root* is balanced of rank k .

Proposition 1.1. *The differential equation*

$$(1) \quad A'_k(x) = 2 \frac{A_k(x)}{1-x} + R'_k(x)$$

holds, with the initial condition $A_k(0) = 0$.

Proof. Note that $a_{n,k}$ is the number of ordered pairs (v, T) , where v is a balanced vertex of rank k in a decreasing binary tree T on vertex set $[n]$, in other words, $a_{n,k}$ is the number of decreasing binary trees on $[n]$ with a balanced vertex of rank k marked. If the marked vertex v is not the root

of T , then removing the root of T , we get, on the one hand, a structure on $[n-1]$ that is counted by $A'_k(x)$, and, on the other hand, a decreasing binary tree and a decreasing binary tree with a balanced vertex of rank k marked. By the Product formula of exponential generating functions, these pairs of trees are counted by the generating function $2 \cdot \frac{1}{1-x} A_k(x)$. The factor 2 is needed since the order of the obtained two trees matters, and $1/(1-x)$ is just the generating function of the sequence of factorials, hence of the sequence enumerating decreasing binary trees. Finally, if the marked vertex v is the root of tree, then removing it we just get a structure enumerated by $R'_k(x)$. \square

The special case of $k = 0$ counts *leaves*, which are of rank 0, and are trivially balanced. Indeed, we have $R_0(x) = x$, so $R'_0(x) = 1$. Therefore (1) reduces to

$$A'_0(x) = 2 \frac{A_0(x)}{1-x} + 1,$$

with $A_0(0) = 0$. The solution of this differential equation is indeed

$$A_0(x) = \frac{1}{3} \cdot \frac{1}{(1-x)^2} + \frac{x}{3} - \frac{1}{3},$$

which is the generating function for the number of all leaves of all decreasing binary trees on n vertices.

If $k = 1$, then $R_1(x) = x^2 + \frac{x^3}{3}$, since both trees on two vertices, and both trees on three vertices in which the root has two children, have a root that is balanced of rank 1. So (1) reduces to

$$A'_1(x) = 2 \frac{A_1(x)}{1-x} + 2x + x^2,$$

with $A_1(0) = 0$. The solution of this initial value problem is

$$A_1(x) = \frac{\frac{1}{3}x^5 - x^3 + x^2}{(1-x)^2}.$$

Further generating functions $A_k(x)$ can theoretically be computed, but one runs out of computing power very fast. A crucial difference from earlier work such as [2] is that for all k , the generating function R_k is a *polynomial function*, since if the root of a tree is balanced and is of rank k , then that tree cannot have more than $2^k - 1$ vertices. We are now going to show that this implies that for all k , the generating function $A_k(x)$ is always a *rational function* of denominator $(1-x)^2$.

Corollary 1.2. *Let k be a fixed nonnegative integer. Let $c_{n,k} = \frac{a_{n,k}}{n! \cdot n}$ be the probability that a vertex chosen uniformly from the set of all $n \cdot n!$ vertices of all decreasing binary trees on $[n]$ is balanced, and is of rank k . Then for any fixed k , the limit*

$$c_k = \lim_{n \rightarrow \infty} c_{n,k}$$

exist. Furthermore, Let $A_k(x) = \frac{P_k(x)}{(1-x)^2}$, that is, let $P_k(x)$ be the numerator of $A_k(x)$. Then $c_k = P_k(1)$.

Note that the fact that the limits c_k exist can also be proved using the techniques of *additive functionals* as explained in [6]. (The number of balanced vertices in a rooted tree is an additive functional.) However, we provide a self-contained proof here.

Proof. Solving the linear differential equation (1) for $A_k(x)$, we get that

$$(2) \quad A_k(x) = \frac{\int (1-x)^2 B'_k(x) dx}{(1-x)^2},$$

where the constant of integration is to be chosen so that the initial condition $A_k(0) = 0$ is satisfied. Recall that $B_k(x)$, and therefore, $B'_k(x)$, is a polynomial function. Therefore, the numerator $P_k(x)$ of the right-hand side of (2) is a polynomial function.

Now let $P_k(x) = P(x)(1-x)^2 + Q(1-x) + R$, where $P(x)$ is a polynomial, and Q and R are complex numbers. Note that $P_k(1) = R$. Then

$$A_k(x) = \frac{P_k(x)}{(1-x)^2} = P(x) + \frac{Q}{1-x} + \frac{R}{(1-x)^2}.$$

If n is larger than the degree of the polynomial P , then equating coefficients of x^n on both sides, we get that

$$\frac{a_{n,k}}{n!} = Q + (n+1)R.$$

So

$$\lim_{n \rightarrow \infty} c_{n,k} = \lim_{n \rightarrow \infty} \frac{a_{n,k}}{n! \cdot n} = \lim_{n \rightarrow \infty} \frac{a_{n,k}}{(n+1)!} = R = P_k(1).$$

□

Note that it is easy to see that $c_k > 0$ for all k . Indeed, c_k is certainly at least as large as the probability that a randomly selected vertex is of rank k and is the root of a *perfect* binary tree (one in which every non-leaf vertex has exactly two children), and even this last probability stays above a positive constant as n goes to infinity. See [1] for details.

The simple form of $A_k(x)$ is the reason that decreasing binary trees are easier to analyze from this aspect than other tree varieties. Using the above corollary, we get that $c_0 = 1/3$, $c_1 = 1/5$, $c_2 = 52/567$ and $c_3 = 7175243/222660900$. This shows that for large n , about 65.7 percent of all vertices of decreasing binary trees are balanced and of rank at most three. More computation shows that for n sufficiently large, about 66.62 percent of all vertices are balanced and of rank at most four, and about 66.84 percent are balanced and of rank at most five.

1.2. A result about monotonicity. As decreasing binary trees are in bijection with permutations, we have additional tools analyzing them. This enables to prove the following strong result. We could not prove similar results for other labeled rooted trees.

Theorem 1.3. *Let P_n be the probability that a vertex chosen uniformly at random from the set of all vertices of all decreasing binary trees on $[n]$ is balanced. Then the sequence P_1, P_2, \dots is weakly decreasing.*

Let $p_{n,k}$ be the probability that the root of a randomly selected tree on n vertices is balanced, and is of rank k . Set $p_{0,i} = 1$ for all i . We start by an inequality for the numbers $p_{n,k}$ for fixed k .

Lemma 1.4. *For all $n \geq 1$ and all fixed $k \leq n$, the inequality $p_{n+1,k} \leq p_{n,k}$ holds.*

Proof. We prove the statement by induction on n . The statement is true for all k if $n \leq 3$, since in that case, $p_{n,k} = 1$ for all n and all k . Now let us assume that the statement is true for n and prove it for $n + 1$.

Let π be a permutation of length $n + 1$. The probability that the largest entry of π is in position $i + 1$ for any $i \in [0, n]$ is $1/(n + 1)$. The root of $T(\pi)$ is balanced of rank k if and only if all its children are balanced of rank $k - 1$, so

$$(3) \quad p_{n+1,k} = \frac{\sum_{i=0}^n p_{i,k-1} p_{n-i,k-1}}{n+1}.$$

Replacing $n + 1$ by n , we get the analogous formula

$$(4) \quad p_{n,k} = \frac{\sum_{i=0}^{n-1} p_{i,k-1} p_{n-1-i,k-1}}{n}.$$

The initial difficulty is that while the summands in (3) are smaller than their counterparts in (4), there is one more of them. The crucial observation is that if we remove the *smallest* summand from the numerator of (3), then the remaining n summands of that numerator can be matched with the n summands of the numerator of (4), so that in each pair, the summand coming from (4) is at least as large as the summand coming from (3). That will prove that the numerator of (3) is at most $(n + 1)/n$ times as large as the numerator of (4).

Let j be the index for which $p_{j,k-1} p_{n-j,k-1}$ is minimal, so that last product is the minimal summand in the numerator of (3). First we look at indices smaller than j . Note that by the induction hypothesis, $p_{n-i,k-1} \leq p_{n-1-i,k-1}$, so

$$(5) \quad p_{i,k-1} p_{n-i,k-1} \leq p_{i,k-1} p_{n-1-i,k-1}.$$

Let us sum these inequalities for $0 \leq i \leq j - 1$, to get

$$(6) \quad \sum_{i=0}^{j-1} p_{i,k-1} p_{n-i-1,k-1} \leq \sum_{i=0}^{j-1} p_{i,k-1} p_{n-1-i,k-1}.$$

Now we consider indices larger than j . Again, by the induction hypothesis, $p_{i,k-1} \leq p_{i-1,k-1}$, so

$$(7) \quad p_{i,k-1} p_{n-i,k-1} \leq p_{i-1,k-1} p_{n-i,k-1}.$$

Let us sum these inequalities for $i \in [j+1, n]$, to get

$$(8) \quad \sum_{i=j+1}^{n-1} p_{i,k-1} p_{n-i,k-1} \leq \sum_{i=j+1}^{n-1} p_{i-1,k-1} p_{n-i,k-1}.$$

Adding (6) and (8), we get an inequality whose right-hand side agrees with the sum in (4), and whose left-hand side is the sum in (3), *except* the summand of the latter indexed by j . However, that summand was the *smallest* of the $(n+1)$ summands in the sum in (3), which implies that

$$np_{n+1,k} = \frac{n}{n+1} \sum_{i=0}^n p_{i,k-1} p_{n-i,k-1} \leq \sum_{i=0}^{n-1} p_{i,k-1} p_{n-1-i,k-1} = np_{n,k}.$$

This is equivalent to our claim. \square

Corollary 1.5. *Let p_n be the probability that the root of a decreasing binary tree on $[n]$ is balanced. Then $p_n \geq p_{n+1}$.*

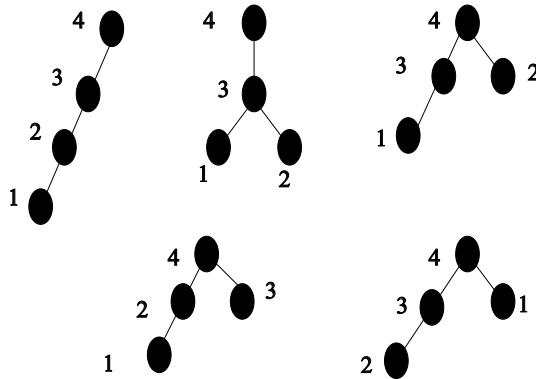
Proof. It follows from our definitions that $p_n = \sum_{k=1}^{n-1} p_{n,k}$ and $p_{n+1} = \sum_{k=1}^n p_{n+1,k}$. As Lemma 1.4 shows that $p_{n+1,k} \leq p_{n,k}$ for $k \leq n$, the only issue that we must consider is that the sum that provides p_{n+1} has one more summand than the sum that provides p_n . However, this is not a problem, since for all $n \geq 2$, we have $p_{n,n-1} = 2^{n-1}/n!$, while $p_{n+1,n-1} = 2^{n-1}/(n+1)!$ and $p_{n+1,n} = 2^n/(n+1)!$, so

$$p_{n,n-1} = \frac{2^{n-1}}{n!} \geq \frac{3 \cdot 2^{n-1}}{n+1} = p_{n+1,n-1} + p_{n+1,n}.$$

This inequality, and applying Lemma 1.4 for all $k \leq n-2$, proves our claim. \square

Proof. (of Theorem 1.3.) Induction on n , the initial case of $n = 1$ being obvious. In order to prove that $P_n \geq P_{n+1}$, note that a random vertex of a tree of size n has $1/n$ probability to be the root, and it has, for each $i \in [n-1]$, exactly $1/n$ probability to be a vertex in a subtree of size i which is the left subtree or right subtree of the root. Therefore, the inequality $P_n \geq P_{n+1}$ is equivalent to the inequality

$$(9) \quad P_n = \frac{p_n + \sum_{i=1}^{n-1} P_i}{n} \geq \frac{p_{n+1} + \sum_{i=1}^n P_i}{n+1} = P_{n+1}.$$

FIGURE 2. The five rooted non-plane 1-2 trees on vertex set $[4]$.

The inequality in (9) is true, since the first equality in (9) shows that P_n is obtained as the average of the n values in the set $S = \{p_n, P_1, P_2, \dots, P_{n-1}\}$. That average does not change if we extend S by adding P_n (the average of the values in S) to it. Then, if we replace p_n by p_{n+1} , the average of the new set $S' = \{p_{n+1}, P_1, P_2, \dots, P_n\}$ is at most as large as the average of S , (since $p_{n+1} \leq p_n$ by Corollary 1.5). The average of S' is P_{n+1} by the same argument that we used to show the leftmost equality in (9), we just need to replace n by $n + 1$. \square

As the sequence P_1, P_2, \dots is weakly decreasing, its limit L exists. Note that $L \geq \sum_{k=0}^5 c_{n,k} \approx 0.6684$ as we mentioned in Section 1.1. On the other hand, the number of *balanced* vertices of rank larger than five is certainly at most as large as the number of *all* vertices of rank larger than five, and it is known [2] that the latter is about 0.00125 times the total number of vertices. This proves that

$$0.6684 \leq L \leq 0.66965.$$

2. NON-PLANE 1-2 TREES

In these trees, the vertices are still bijectively labeled by the elements of $[n]$, each vertex has a smaller label than its parent, and each non-leaf vertex has one or two children, but "left" and "right" do not matter anymore. See Figure 2 for an illustration. It is well known [3] that the number of such trees is the Euler number E_n , and that the exponential generating function of the Euler numbers is

$$y(x) = \sum_{n \geq 0} E_n \frac{x^n}{n!} = \tan x + \sec x.$$

See sequence A000111 in the *On-line Encyclopedia of Integer Sequences* [7] for the many occurrences of these numbers in Combinatorics.

Let $\mathcal{A}_k(x)$ be the exponential generating function for the number of all balanced vertices of rank k in all non-plane 1-2 trees on $[n]$, and let $\mathcal{R}_k(x)$ be the exponential generating function for the number of non-plane 1-2 trees on $[n]$ in which *the root* is balanced and of rank k .

Theorem 2.1. *The differential equation*

$$(10) \quad \mathcal{A}'_k(x) - \mathcal{A}_k(x)y(x) = \mathcal{R}'_k(x)$$

holds, with the initial condition $\mathcal{A}_k(0) = 0$.

Proof. Let (v, T) be an ordered pair in which T is a non-plane 1-2 tree on vertex set $[n]$ and v is a balanced vertex of rank k of T . Then $\mathcal{A}_k(x)$ is the exponential generating function counting such pairs. Let us first assume that v is not the root of T , and let us remove the root of T . On the one hand, this leaves a structure that is counted by $\mathcal{A}'_k(x)$. On the other hand, this leaves an ordered pair consisting of a non-plane 1-2 tree with a vertex of order k marked, and a non-plane 1-2 tree. By the Product formula of exponential generating functions, such ordered pairs are counted by the generating function $\mathcal{A}_k(x)y(x)$. Finally, v was the root of T , then the root of T was balanced and of rank k . Such trees are counted by $\mathcal{R}_k(x)$, or, after the removal of their root, by $\mathcal{R}'_k(x)$. \square

Crucially, the generating function $\mathcal{R}_k(x)$, and hence, its derivative $\mathcal{R}'_k(x)$, are *polynomials*, which enables us to explicitly solve the linear differential equation (10). Indeed, the solution is

$$(11) \quad \mathcal{A}_k(x) = \frac{\int \mathcal{R}'_k(x)(1 - \sin x) dx}{1 - \sin x},$$

where the integral in the numerator is an elementary function since the integral of $x^n \sin x$ is an elementary function for all positive integers n . (The constant of integration is chosen so that the initial condition $\mathcal{A}_k(0) = 0$ is satisfied.) Note that we are not able to count *all* vertices of rank k in a similar fashion, since the generating function for the number of non-plane 1-2 trees in which the root is of rank k is not a polynomial, and the solutions analogous to (11) will not be elementary functions for $k \geq 1$. Even $\int x \tan x dx$ and $\int x \sec x dx$ are not elementary functions.

2.1. The number of all vertices. So that we could compute the probability that a randomly selected vertex of a randomly selected non-plane 1-2 tree on $[n]$ is balanced and of rank k , we need to know the size of the number nE_n of *all* vertices in all such trees. The asymptotics of the Euler numbers are well-known (see [5], for example), but to keep the paper self-contained, we provide an argument here at the level of precision that we will need. See any introductory textbook on Complex analysis, such as [4] for the relevant notions.

As $y(x) = \tan x + \sec x$, the dominant singularities of $y(x)$ are at $x = \pi/2$ and $x = -\pi/2$, so the coefficients of $y(x)$ are of exponential order $2/\pi$. (Note

that the singularity at $x = -\pi/2$ is removable.) However, we need a little more more precision. The following proposition will provide that.

Proposition 2.2. *Let $H(x) = f(x)/g(x)$ be a function so that $f(x)$ and $g(x)$ are analytic functions, $f(x_0) \neq 0$, while $g(x) = 0$ and $g'(x) \neq 0$. Then*

$$\text{Res } H(x) \Big|_{x_0} = \frac{f(x_0)}{g'(x_0)}.$$

We can apply Proposition 2.2 to $y(x)$ at $x_0 = \pi/2$ with $f(x) = 1 + \sin x$ and $g(x) = \cos x$ if we note that $y(x) = \frac{1+\sin x}{\cos x}$. Then Proposition 2.2 implies that $\text{Res } y(x) \Big|_{\pi/2} = \frac{2}{-1} = -2$. The singularity of y at $x = -\pi/2$ is removable, since $\lim_{x \rightarrow -\pi/2} y(x) = 0$ exists, so $\text{Res } y(x) \Big|_{-\pi/2} = 0$.

Now observe that

$$(12) \quad \frac{R}{x-a} = \frac{R}{-a} \cdot \frac{1}{1 - \frac{x}{a}}$$

$$(13) \quad = \frac{R}{-a} \sum_{n \geq 0} \frac{x^n}{a^n}.$$

Applying this to $y(x)$ with $a = \pi/2$ and $R = -2$, we get that the dominant term of $y(x)$ is of the form $\frac{4}{\pi} \sum_{n \geq 0} x^n (2/\pi)^n$, so

$$(14) \quad \frac{E_n}{n!} \sim \frac{4}{\pi} \cdot \left(\frac{2}{\pi}\right)^n.$$

So the total number of all vertices in all non-plane 1-2 trees of size n is

$$(15) \quad nE_n \sim n! \left(n \frac{4}{\pi} \cdot \left(\frac{2}{\pi}\right)^n \right).$$

2.2. Leaves. Let $\mathcal{A}_0(x)$ denote exponential generating function for the total number of *leaves* in all non-plane 1-2 trees on vertex set $[n]$. It is then easy to verify that $\mathcal{A}_0(x) = x + \frac{x^2}{2} + 2\frac{x^3}{6} + 9\frac{x^4}{24} + \dots$.

Theorem 2.3. *The equality*

$$(16) \quad \mathcal{A}_0(x) = \frac{x - 1 + \cos x}{1 - \sin x}$$

holds.

Proof. Let us apply Theorem 2.1 with $k = 0$, to get the linear differential equation

$$\mathcal{A}'_0(x) - \mathcal{A}_0(x)y(x) = 1.$$

Indeed, $R_0(x) = x$, since the only tree whose root is balanced and of rank 0 is the one-vertex tree. Recalling that $y(x) = \tan x + \sec x$, and the initial condition $\mathcal{A}_0(0) = 0$, we can solve the last displayed linear differential equation to get what was to be proved. \square

Note that $x = \pi/2$ is the unique nonremovable singularity of smallest modulus of $\mathcal{A}_0(x)$, and that at that point, $\mathcal{A}_0(x)$ has pole of order two, since $(1 - \sin x)' = -\cos x$ also has a zero at that point. Therefore, we cannot apply Proposition 2.2 directly. Instead, we use the following lemma.

Lemma 2.4. *Let $H(x) = \frac{f(x)}{g(x)}$ be a function so that f and g are analytic functions, $f(x_0) \neq 0$, while $g(x_0) = g'(x_0) = 0$, and $g''(x) \neq 0$. Then*

$$H(x) = \frac{2f(x_0)}{g''(x_0)} \cdot \frac{1}{(x - x_0)^2} + \frac{h_{-1}}{x - x_0} + h_0 + \dots$$

Proof. The conditions directly imply that g has a double root, and hence H has a pole of order two, at x_0 . In order to find the coefficient that belongs to that pole, let $g(x) = q(x)(x - x_0)^2$. Now differentiate both sides with respect to x , to get

$$g''(x) = q''(x)(x - x_0)^2 + 4q'(x)(x - x_0) + 2q(x).$$

Setting $x = x_0$, we get

$$(17) \quad g''(x_0) = 2q(x_0).$$

By our definitions, in a neighborhood of x_0 , the function $H(x)$ behaves like

$$\frac{f(x)}{q(x)(x - x_0)^2},$$

and our claim follows by (17). \square

Let $\mathcal{A}_{n,0}$ be the number of all leaves in all non-plane 1-2 trees on vertex set $[n]$.

Theorem 2.5. *The equality*

$$\mathcal{C}_0 := \lim_{n \rightarrow \infty} \frac{\mathcal{A}_{n,0}}{nE_n} = 1 - \frac{2}{\pi} \approx 0.3633802278$$

holds. In other words, for large n , the probability that a vertex chosen uniformly at random from all vertices of all non-plane 1-2 trees is a leaf is $1 - \frac{2}{\pi}$.

Proof. Note that $\mathcal{A}_0(x)$ has a unique singularity of smallest modulus, at $x = \pi/2$, so the exponential growth rate of its coefficients is $2/\pi$. Also note that at that point, the denominator of $\mathcal{A}_0(x)$ has a double root. Therefore, Lemma 2.4 applies, with $f(x) = x - 1 + \cos x$ and $g(x) = 1 - \sin x$, yielding that the coefficient of the $(x - \pi/2)^{-2}$ term in the Laurent series of $\mathcal{A}_0(x)$ about $x_0 = \pi/2$ is

$$2 \cdot \frac{(\pi)/2 - 1 + \cos(\pi/2)}{\sin(\pi/2)} = \pi - 2.$$

Now observe that

$$(18) \quad \frac{D}{(x - a)^2} = \frac{D}{a^2} \cdot \frac{1}{(1 - \frac{x}{a})^2} = \frac{D}{a^2} \cdot \sum_{n \geq 0} (n + 1) \frac{x^n}{a^n}.$$

Applying this to the dominant term of $\mathcal{A}_0(x)$ with $D = \pi - 2$ and $a = \pi/2$, we get that

$$(19) \quad \frac{\mathcal{A}_{n,0}}{n!} \sim n(\pi - 2) \cdot \left(\frac{2}{\pi}\right)^{n+2}.$$

The proof of our claim is now immediate by comparing formulas (19) and (15). \square

2.3. Balanced vertices of rank 1. Let $\mathcal{A}_1(x)$ be exponential generating function for the total number of balanced vertices of rank 1 in all non-plane 1-2 trees on vertex set $[n]$. Note that such vertices have only leaves as neighbors.

Theorem 2.6. *The differential equation*

$$(20) \quad \mathcal{A}_1(x) = \frac{1}{6} \cdot \frac{(3x^2 + 6x - 6) \cos x - (6x + 6) \sin x + x^3 + 3x^2 + 6}{1 - \sin x}$$

holds, with the initial condition $\mathcal{A}_1(0) = 0$.

Proof. Let us apply Theorem 2.1 with $k = 1$, to get the linear differential equation

$$\mathcal{A}'_1(x) - \mathcal{A}_1(x)y(x) = x + \frac{x^2}{2}.$$

Indeed, $R_1(x) = \frac{x^2}{2} + \frac{x^3}{6}$, since there are only two trees whose root is balanced and of rank 1; one of them has two vertices and the other one has three vertices. Recalling that $y(x) = \tan x + \sec x$, and the initial condition $\mathcal{A}_1(0) = 0$, we can solve the last displayed differential equation to prove our claim. \square

Theorem 2.7. *The equality*

$$\mathcal{C}_1 := \lim_{n \rightarrow \infty} \frac{\mathcal{A}_{n,1}}{nE_n} = \frac{\pi^2}{24} + \frac{\pi}{4} - 1 \sim 0.1966316804$$

holds. In other words, for large n , the probability that a vertex chosen uniformly at random from all vertices of all non-plane 1-2 trees is balanced and of rank 1 is $\frac{\pi^2}{24} + \frac{\pi}{4} - 1$.

Proof. Note that $\mathcal{A}_1(x)$ has a unique singularity of smallest modulus at $x = \pi/2$, and that that singularity is a pole of order two. Therefore, we can apply Lemma 2.4 with $f(x) = (3x^2 + 6x - 6) \cos x - (6x + 6) \sin x + x^3 + 3x^2 + 6$ and $g(x) = 6(1 - \sin x)$. At $x_0 = \pi/2$, this yields $f(x_0) = \frac{\pi^3}{8} + \frac{3\pi^2}{4} - 3\pi$. Furthermore, $g''(x) = 6 \sin x$, so $g''(x_0) = 6$. Therefore, the coefficient of the $1/(x - \pi/2)^2$ term in the Laurent series of $\mathcal{A}_1(x)$ about $x_0 = \pi/2$ is

$$\frac{2f(x_0)}{g''(x_0)} = \frac{2}{6} \cdot \left(\frac{\pi^3}{8} + \frac{3\pi^2}{4} - 3\pi \right) = \frac{\pi^3}{24} + \frac{\pi^2}{4} - \pi.$$

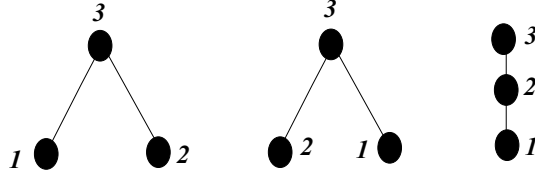


FIGURE 3. The three rooted plane 1-2 trees on vertex set [3].

Applying (18) with $D = \frac{\pi^3}{24} + \frac{\pi^2}{4} - \pi$ and $a = \pi/2$, we get that

$$(21) \quad \frac{\mathcal{A}_{n,1}}{n!} \sim n \left(\frac{\pi^3}{24} + \frac{\pi^2}{4} - \pi \right) \cdot \left(\frac{2}{\pi} \right)^{n+2}.$$

We can now prove our claim by comparing formulae (15) and (21). \square

2.4. Balanced vertices of higher rank. For any fixed k , we can compute the probability that a vertex selected from all vertices of all non-plane 1-2 trees uniformly at random is balanced and of rank k . For instance, for $k = 2$, we get that

$$\lim_{n \rightarrow \infty} \frac{\mathcal{A}_{n,2}}{nE_n} = \frac{\pi^6}{32256} + \frac{\pi^5}{2304} - 7 \frac{\pi^4}{1920} - 3 \frac{\pi^3}{64} + \frac{\pi^2}{3} + 9 \frac{\pi}{4} - 8 - \frac{2}{\pi} \approx 0.0759013197,$$

where $\mathcal{A}_{n,2}$ denotes the number of balanced vertices in all non-plane 1-2 trees on $[n]$.

It is a direct consequence of (11), the well-known fact that (easy to prove by induction) that

$$\int x^n \sin x \, dx = H(x) \sin x + I(x) \cos x,$$

for some polynomials $H(x)$ and $I(x)$, and Lemma 2.4 that for all positive integers n , there exists a polynomial J_n with rational coefficients so that

$$C_k := \lim_{n \rightarrow \infty} \frac{\mathcal{A}_{n,k}}{nE_n} = \frac{J_n(\pi)}{\pi}.$$

3. PLANE 1-2 TREES

Plane 1-2 trees are similar to non-plane 1-2 trees, except that the children of each vertex are linearly ordered, left to right. The difference between plane 1-2 trees and decreasing binary trees is that in plane 1-2 trees, if a vertex has only one child, then that child has no "direction", that is, it is not a "left child" or a "right child". See Figure 3 for an illustration.

So plane 1-2 trees are "in between" decreasing binary trees (where left or right matters for every child) and non-plane 1-2 trees (where left or right does not matter for any vertex). Indeed, in plane 1-2 trees, left or right matters, except for vertices that have no siblings.

Let $Z(x) = \sum_{n \geq 0} z_n \frac{x^n}{n!}$ be the exponential generating function for the number of plane 1-2 trees on vertex set $[n]$. Then $Z(x)$ satisfies the differential equation

$$(22) \quad Z'(x) = Z^2(x) - Z(x) + 1,$$

with initial condition $Z(0) = 1$. Indeed, removing the root of a plane 1-2 tree T that has more than one vertex, that tree falls apart to the ordered set of two such trees, except when the root of T has only one child.

See sequence A080635 in [7] for many other combinatorial problems whose solution involves the power series $Z(x)$.

Solving (22), we get the explicit formula

$$Z(x) = \frac{\sqrt{3}}{2} \tan \left(\frac{\sqrt{3}}{2} x + \frac{\pi}{6} \right) + \frac{1}{2}.$$

Noting that $\tan a = \sin a / \cos a$, and that the summand $1/2$ at the end does not influence the growth rate of the coefficients of $Z(x)$, we can proceed as in Section 2. That is, we can use Proposition 2.2 to compute that the number z_n of plane 1-2 trees on vertex set $[n]$ satisfies

$$z_n \sim n! \left(\frac{3\sqrt{3}}{2\pi} \right)^{n+1}.$$

Therefore, the total number of all vertices of all such trees satisfies

$$(23) \quad nz_n \sim n! \cdot n \left(\frac{3\sqrt{3}}{2\pi} \right)^{n+1}.$$

Let $\mathbf{A}_k(x)$ be the exponential generating function for the number of all balanced vertices of rank k in all plane 1-2 trees on $[n]$, and let $\mathbf{R}_k(x)$ be the exponential generating function for the number of plane 1-2 trees on $[n]$ in which *the root* is balanced and of rank k .

Theorem 3.1. *The differential equation*

$$(24) \quad \mathbf{A}'_k(x) = 2\mathbf{A}_k(x)Z(x) - \mathbf{A}_k(x) + \mathbf{R}'_k(x)$$

holds, with the initial condition $\mathbf{A}_k(0) = 0$.

Proof. Let (v, T) be an ordered pair in which T is a plane 1-2 tree on vertex set $[n]$ and v is a balanced vertex of rank k of T . Then $\mathbf{A}_k(x)$ is the exponential generating function counting such pairs. Let us first assume that v is not the root of T , and let us remove the root of T . On the one hand, this leaves a structure that is counted by $\mathbf{A}'_k(x)$. On the other hand, this leaves an order pair consisting of a plane 1-2 tree with a balanced vertex of rank k marked, and a plane 1-2 tree. If the tree without the marked vertex was not empty, then by the Product formula of exponential generating functions, such ordered pairs are counted by the generating function $2\mathbf{A}_k(x)Z(x)$, since the order of the two trees matters. If the tree without the marked vertex was

empty, then there is only one way to "order" the 1-element set of subtrees of the root, consisting of the subtree with the marked vertex. This results in the correction term $-\mathbf{A}'_k(x)$.

Finally, v was the root of T , then the root of T was balanced and of rank k . Such trees are counted by $\mathbf{R}_k(x)$, or, after the removal of their root, by $\mathbf{R}'_k(x)$. \square

3.1. Leaves. Setting $k = 0$ in (24), noting that $\mathbf{A}_0(0) = 0$, and $\mathbf{R}'_0(x) = 1$, then solving the resulting differential equation we get the following result for the number of all leaves.

Corollary 3.2. *The equality*

$$(25) \quad \mathbf{A}_0(x) = \frac{\frac{\sqrt{3}}{6} \sin\left(\sqrt{3}x + \frac{\pi}{3}\right) + \frac{x}{2} - \frac{1}{4}}{\cos^2\left(\frac{\sqrt{3}x}{2} + \frac{\pi}{6}\right)}$$

holds.

We can apply Lemma 2.4 to the numerator and the denominator of $\mathbf{A}_0(x)$ in (25) to compute the growth rate of the coefficients of that power series. Indeed, $\mathbf{A}_0(x)$ has a unique singularity of smallest modulus at $x_0 = 2\pi/(3\sqrt{3})$, and that point the numerator of $\mathbf{A}_0(x)$ is nonzero, the denominator, and its first derivative are zero, while the second derivative of the denominator is not 0. Then a computation analogous to that immediately following the proof of Theorem 2.7 shows that if $\mathbf{a}_{n,0}$ is the number of all leaves in all plane 1-2 trees on vertex set $[n]$, then

$$(26) \quad \mathbf{a}_{n,0} \sim \left(-\frac{1}{3} + \frac{4}{27}\sqrt{3}\pi\right) \cdot n \left(\frac{3\sqrt{3}}{2\pi}\right)^n.$$

Comparing (26) with (23), we obtain the following result.

Theorem 3.3. *The equality*

$$\mathbf{C}_0 := \lim_{n \rightarrow \infty} \frac{\mathbf{a}_{n,0}}{nz_n} = \frac{2}{3} - \frac{\sqrt{3}}{2\pi} \approx 0.391002219$$

holds.

In other words, the probability that a vertex selected uniformly at random from all plane 1-2 trees of size n will be a leaf is about 0.391.

3.2. Vertices of higher rank. Note that remarkably, we can obtain an explicit formula for $\mathbf{A}_k(x)$ for every k . This is in contrast to the case when we want to compute the generating function of all vertices of a given rank, where we run into non-elementary functions for $k \geq 2$.

Indeed, the standard form of (24) is

$$\mathbf{A}'_k(x) + (1 - 2Z(x))\mathbf{A}_k(x) = \mathbf{R}'_k(x).$$

In order to solve this linear differential equation, first we multiply both sides by the integrating factor

$$\begin{aligned}\mu(x) &= \exp\left(\int(1-2Z(x))dx\right) = \exp\left(-\ln\left(\sec^2\left(\frac{\sqrt{3}x}{2} + \frac{\pi}{6}\right)\right)\right) \\ &= \cos^2\left(\frac{\sqrt{3}x}{2} + \frac{\pi}{6}\right),\end{aligned}$$

which is an elementary function.

After that multiplication, we need to integrate both sides of the obtained equation to get $\mathbf{A}_k(x)$. However, we are able to do so, since on the right hand side, we have $\mu(x)\mathbf{R}'_k(x)$, that is, a cosine function times a *polynomial* function, and it is well known that such products have elementary integrals.

4. FURTHER DIRECTIONS

The method that we used in this paper may well be applicable to count balanced vertices in other tree varieties, as long as the number of children each vertex can have is *bounded*.

It seems intuitively very likely that in any tree variety, as n goes to infinity, the probability that a vertex chosen uniformly at random from all vertices of all trees of size n is balanced will be monotone decreasing. Still in the one case where we could prove this, the case of decreasing binary trees, our proof heavily depended on the simple bijection between these trees and permutations. New ideas are needed for other tree varieties.

We mentioned that plane 1-2 trees are "in between" the other two tree varieties studied in this paper. So it is perhaps interesting that their vertices are *the most likely* to be leafs. Indeed, a random vertex of a decreasing binary tree has a one-third chance to be a leaf, while the same probability is about 36.3 percent for non-plane 1-2 trees, and 39.1 percent for plane 1-2 trees. Understanding this phenomenon could lead to new insights.

REFERENCES

- [1] M. Bóna k -protected vertices in binary search trees. *Adv. in Appl. Math.* **53** (2014), 1–11.
- [2] M. Bóna, B. Pittel, On a random search tree: asymptotic enumeration of vertices by distance from leaves. *J. Appl. Prob.* to appear. Preprint available at at arXiv:1412.2796.
- [3] BÓNÁ, M. (2016) Introduction to Enumerative and Analytic Combinatorics, CRC Press, 2016.
- [4] J. Brown, R. V. Churchill, Complex Variables and Applications, 9th edition. McGraw-Hill, 2013.
- [5] P. Flajolet and R. Sedgewick, Analytic Combinatorics, Cambridge University Press, Cambridge, UK, 2009.
- [6] S. Janson and C. Holmgren, Limit laws for functions of fringe trees for binary search trees and recursive trees. *Electronic J. Probability* 20 (2015), no. 4, 1-51.
- [7] Online Encyclopedia of Integer Sequences, *online database*, www.oeis.org.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF FLORIDA, 358 LITTLE HALL, PO
Box 118105, GAINESVILLE, FL, 32611 – 8105 (USA)

E-mail address: `bona@ufl.edu`